

вИДержавний вищий навчальний заклад
«Донецький національний технічний університет»
Кафедра прикладної математики та інформатики

«ЗАТВЕРДЖУЮ»

Перший проректор

_____ Леонід Бачурін

«_____» _____ 2024 р.

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

ДВС 2.04 «Сучасні методи обробки природної мови»

(шифр і назва навчальної дисципліни)

Рівень освіти: **другий магістерський**

Спеціальність **122 Комп'ютерні науки**

Освітня програма **Комп'ютерні науки**

Мова навчання: **українська**

Луцьк – 2024

Робоча програма навчальної дисципліни «Сучасні методи обробки природної мови» для здобувачів вищої освіти за спеціальністю 122 Комп'ютерні науки.

«27» грудня 2023 р. – 7 с.

Розробник:

Башков Є.О., д.т.н., професор кафедри ПМІ, професор

Робоча програма затверджена на засіданні кафедри **прикладної математики та інформатики**
Протокол № 13 від «27» грудня 2023 р.

Завідувач кафедри прикладної математики та інформатики ДВНЗ «ДонНТУ»

(Маслова Н.О.)
(підпис) (прізвище та ініціали)

«27» грудня 2023 р.

Схвалено науково-методичною комісією з галузі знань 12 «Інформаційні технології»
Протокол № 1 від «15» січня 2024 р.

Голова (Башков Є.О.)
(підпис) (прізвище та ініціали)

«15» січня 2024 р.

1. Загальна інформація

Форма навчання	Денна	Заочна
Статус	Вибіркова	
Обсяг в кредитах ЄКТС	7	
Обсяг в годинах за навчальним планом, разом: в тому числі:	210	
лекції:	48	8
практичні заняття:	–	
лабораторні заняття:	48	8
семінари:	–	
самостійна робота:	114	194
Форма підсумкового контролю	Іспит	
Індивідуальне завдання	Курсова робота	
Дисципліну викладають	Викладач – Башков Євген Олександрович, https://donntu.edu.ua/knt/pmi , yevhen.bashkov@donntu.edu.ua Асистент Нікітенко Андрій Олександрович, https://donntu.edu.ua/kitaer/pmi andrii.nikitenko@donntu.edu.ua	

Передумови для вивчення дисципліни: успішному вивченню дисципліни «Сучасні методи обробки природної мови» сприяє попереднє опанування такими дисциплінами, як «Основи алгоритмізації», «Математичні методи дослідження операцій», «Теорія ймовірностей і математична статистика», «Емпіричні методи програмної інженерії».

2. Мета вивчення навчальної дисципліни «Сучасні методи обробки природної мови»

Дисципліна «Сучасні методи обробки природної мови» має на меті надання цілісного представлення здобувачам вищої освіти сучасних методів та алгоритмів обробки та генерації текстової інформації при розробці та дослідженні систем відтворення природної мови.

Компетентності:

ЗК01. Здатність до абстрактного мислення, аналізу та синтезу.

ЗК02. Здатність застосовувати знання у практичних ситуаціях.

ЗК03. Здатність спілкуватися державною мовою як усно, так і письмово.

ЗК04. Здатність спілкуватися іноземною мовою.

ЗК05. Здатність вчитися й оволодівати сучасними знаннями.

ЗК06. Здатність бути критичним і самокритичним.

ЗК07. Здатність генерувати нові ідеї (креативність).

ФК01. Усвідомлення теоретичних засад комп'ютерних наук.

ФК05. Здатність розробляти, описувати, аналізувати та оптимізувати архітектурні рішення інформаційних та комп'ютерних систем різного призначення.

ФК06. Здатність застосовувати існуючі і розробляти нові алгоритми розв'язування задач у галузі комп'ютерних наук.

ФК11. Здатність ініціювати, планувати та реалізовувати процеси розробки інформаційних та комп'ютерних систем та програмного забезпечення, включно з його

розробкою, аналізом, тестуванням, системною інтеграцією, впровадженням і супроводом.

ФК9. Здатність до розроблення нових та вдосконалення існуючих моделей, методів, засобів, процесів у сфері програмної інженерії, які забезпечують розвиток або надають нові можливості технологій розробки та використання програмного забезпечення.

Програмні результати навчання:

ПРН01. Мати спеціалізовані концептуальні знання, що включають сучасні наукові здобутки у сфері комп'ютерних наук і є основою для оригінального мислення та проведення досліджень, критичне осмислення проблем у сфері комп'ютерних наук та на межі галузей знань.

ПРН02. Мати спеціалізовані уміння/навички розв'язання проблем комп'ютерних наук, необхідні для проведення досліджень та/або провадження інноваційної діяльності з метою розвитку нових знань та процедур.

ПРН04. Управляти робочими процесами у сфері інформаційних технологій, які є складними, непередбачуваними та потребують нових стратегічних підходів.

ПРН05. Оцінювати результати діяльності команд та колективів у сфері інформаційних технологій, забезпечувати ефективність їх діяльності.

ПРН08. Розробляти математичні моделі та методи аналізу даних (включно з великим).

ПРН09. Розробляти алгоритмічне та програмне забезпечення для аналізу даних (включно з великими).

ПРН10. Проектувати архітектурні рішення інформаційних та комп'ютерних систем різного призначення

ПРН11. Створювати нові алгоритми розв'язування задач у сфері комп'ютерних наук, оцінювати їх ефективність та обмеження на їх застосування.

ПРН14. Тестувати програмне забезпечення.

ПРН15. Виявляти потреби потенційних замовників щодо автоматизації обробки інформації.

ПРН16. Виконувати дослідження у сфері комп'ютерних наук.

ПРН17. Виявляти та усувати проблемні ситуації в процесі експлуатації програмного забезпечення, формулювати завдання для його модифікації або реінжинірингу

ПРН19. Аналізувати сучасний стан і світові тенденції розвитку комп'ютерних наук та інформаційних технологій

3. Очікувані результати навчання

Основними результатами опанування дисципліни «Сучасні методи обробки природної мови» є:

- Розвиток творчих здібностей щодо прогнозування напрямків та шляхів розвитку науки та практики побудови інтелектуальних інформаційних систем з використанням підходів штучних нейронних мереж та глибокого навчання.

- Глибоке та творче засвоєння теоретичних знань щодо аналізу, оцінювання, розробки та використання сучасних методів побудови інтелектуальних інформаційних систем на базі глибокого навчання.

- Стимулювання здобувачів до інноваційної діяльності у галузі інформаційних технологій.

- Сприяння теоретичній підготовці здобувачів для вирішення завдань дисертаційних досліджень у галузі інформаційних технологій з використання методів глибокого навчання.

- Активізація творчого відношення до професійної діяльності

- Застосовування отриманих теоретичних знань на практиці та аналіз отриманих результатів.

Внаслідок вивчення курсу студенти повинні вміти:

- Застосовувати знання для постановки і вирішення актуальних наукових завдань, обґрунтування, та використання відповідних методів глибокого навчання для побудови інтелектуальних інформаційних систем.
- Прогнозувати та оцінювати ефективність методів глибокого навчання для побудови інтелектуальних інформаційних систем.
- Проектувати та впроваджувати інтелектуальні інформаційні системи, побудованих на базі методів глибокого навчання.
- Застосовувати принципи навчання впродовж життя для вирішення завдань поза програмою курсу.

4. Засоби діагностики результатів навчання

Засобами оцінювання та методами демонстрування результатів навчання при опануванні дисципліною «Сучасні методи обробки природної мови»:

- екзамен;
- стандартизовані тести;
- виконання завдань на лабораторному обладнанні.

5. Критерії оцінювання результатів навчання

Максимальний бал, визначений схемою оцінювання, наведеною нижче, можливо отримати за умови своєчасного та правильного виконання завдань.

За наявності помилок або при несвоєчасному виконанні оцінка знижується до 60% від максимальної.

Поточний контроль для денної форми навчання						Поточний контроль	Екзамен	Максимальна сума балів
Лр 1	Лр 2	Лр 3	Лр 4	Лр 5	Лр 6			
4	6	6	8	8	8	40	60	100
2	4	4	4	4	5	24		

Примітки: 1) Лр1, Лр2 і т.д лабораторні роботи;
2) У числівнику максимальний бал – при своєчасному та правильному виконанні, у знаменнику – мінімальний (при правильному, але несвоєчасному виконанні)

Поточний контроль для заочної форми навчання				Поточний контроль	Екзамен	Максимальна сума балів
Лр 1	Лр 2	Лр 3	Лр 4			
8	10	10	12	40	60	100
4	6	6	8	24		

Примітки: 1) Пр1, Пр2 і т.д практичні заняття;
2) У чисельнику максимальний бал – при своєчасному та правильному виконанні, у знаменнику – мінімальний (при правильному, але несвоєчасному виконанні)

Оцінювання знань студента при здачі іспиту здійснюється за 100 бальною шкалою.

Контроль виконання курсового проекту включає поточний контроль за виконанням розрахунків та захист перед комісією. Оцінка виконання та захисту курсової роботи проводиться за 100-бальною шкалою.

Розподіл балів, які отримують студенти за виконання курсового проекту

Пояснювальна записка	Захист роботи	Сума
----------------------	---------------	------

до 40	до 60	100
-------	-------	-----

Оцінювання знань студента при здачі іспиту та результатами поточної роботи здійснюється за 100 бальною шкалою.

Оцінка	
За 100-бальною шкалою	Для екзамену, курсового проекту(роботи), практики, диференційованого заліку, кваліфікаційного екзамену, випускної кваліфікаційної (дипломної) роботи (проекту)
90-100	відмінно
74-89	добре
60-73	задовільно
0-59	незадовільно

6. Програма навчальної дисципліни

6.1. Основні теми дисципліни

Тема 1. Вступ до глибокого навчання. Визначення нейронних мереж та глибокого навчання. Основи нейронних мереж. Типи нейронних мереж. Алгоритм зворотного поширювання.

Тема 2. Основи обробки природної мови (NLP). Завдання та застосування NLP. Попередня обробка тексту. Представлення тексту. One-Hot. Word Embedding.

Тема 3. Моделі NLP. Рекурентні нейронні мережі (RNN). Мережі з довгостроковою короткочасною пам'яттю (LSTM). Гейтовані рекурентні блоки (GRU). Моделі послідовності (Seq2Seq). Порівняння моделей.

Тема 4. Згорткові нейронні мережі (CNN) та NLP. Класифікація тексту за допомоги CNN. Класифікація речень за допомоги CNN. Розпізнавання іменованих сутностей / об'єктів.

Тема 5. Трансформери. Моделі трансформерів та механізм уваги. Трансформерне навчання. Передача навчання.

Тема 6. Передові результати NLP. Генеративні моделі. Передача навчання для завдань NLP. Тонке налаштування попередньо навчених моделей для завдань НЛП

6.2. Теми практичних занять

Не передбачено навчальним планом

6.3. Теми лабораторних занять

№ з/п	Назва теми	Кількість годин, денна форма	Кількість годин, заочна форма
1	Лабораторна робота № 1. (Теми 2) Попередня обробка BERT з текстом TF	6	2
2	Лабораторна робота № 2. (Тема 2) Токенізатори підслів	6	2
3	Лабораторна робота № 3. (Тема 2) Класифікація тексту за допомогою BERT	9	3
4	Лабораторна робота № 4. (Тема 3) Класифікація тексту за допомогою RNN	9	3
5	Лабораторна робота № 5. (Тема 5) Модель-трансформер	9	

	для розуміння мови		
6	Лабораторна робота. (Тема 5) Нейронний машинний переклад з увагою seq2seq	9	
	Разом	48	8

6.4. Самостійна робота

№ з/п	Назва теми	Кількість годин, денна форма	Кількість годин, заочна форма
1	Тема 1. Вступ до глибокого навчання	20	30
2	Тема 2. Основи обробки природної мови (NLP)	20	35
3	Тема 3. Моделі NLP	20	35
4	Тема 4. Згорткові нейронні мережі (CNN) та NLP.	20	35
5	Тема 5. Трансформери та механізм уваги	20	35
6	Тема 6. Передові результати NLP	14	24
	Разом	114	194

6.5. Індивідуальне завдання

Курсова робота

Загальна тематика курсової роботи: «Системи класифікації текстової інформації з використанням нейронних мереж глибокого навчання».

7. Література

7.1. Основна

1. Sarkar D. Text Analytics with Python - Bangalore, Karnataka, India, 2019. – 674 p.
2. Goldberg Y. Neural Network Methods for Natural Language Processing / Morgan & Claypool Publ, 2017. – 285 p.
3. Bengfort B., Ojeda T., Bilbro R. Applied Text Analysis with Python. Enabling Language Aware Data Products with Machine Learning / O'Reilly Media, 2016. – 362 p.

7.2. Допоміжна

1. Tunstall L., von Werra L., Wolf T. Natural Language Processing with Transformers / O'Reilly Media, 2022. – 297 p.
2. Jurafsky D., Martin J.H. Speech and Language Processing An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition / Stanford University, 2023 – 626 p.

8. Інформаційні ресурси

1. Обробка природної мови. Режим доступу: https://uk.wikipedia.org/wiki/%D0%9E%D0%B1%D1%80%D0%BE%D0%B1%D0%BA%D0%B0_%D0%BF%D1%80%D0%B8%D1%80%D0%BE%D0%B4%D0%BD%D0%BE%D1%97_%D0%BC%D0%BE%D0%B2%D0%B8
2. Обробка природної мови (NLP) Режим доступу: <https://metinvest.digital/ru/page/1052>